Advanced Topics in Machine Learning

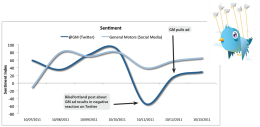
A. LAZARIC (*INRIA-Lille*)

DEI, Politecnico di Milano

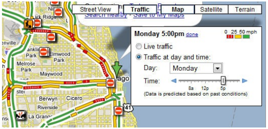
SequeL – INRIA Lille

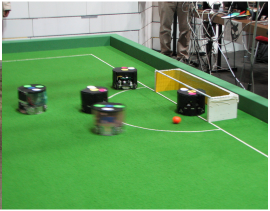
April 2-15, 2012

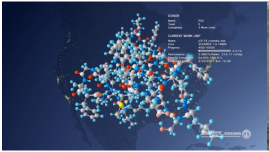
Before You Attended the Course



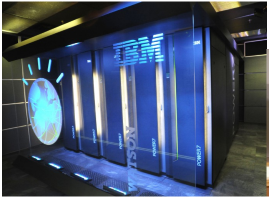
Online @ds

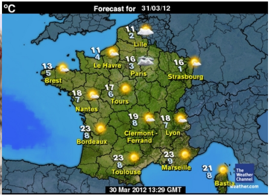


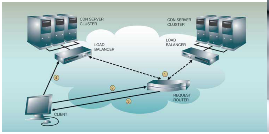


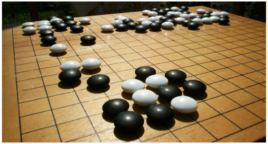


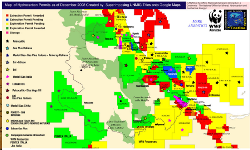
WolframAlpha™
computational™ knowledge engine




°C
Forecast for 31/03/12
Lille
Le Havre
Paris
Strasbourg
Brest
Tours
Nantes
Clermont - Ferrand
Lyon
Bordeaux
Toulouse
Marseille
Bastia
The Weather Channel
weather.com
30 Mar 2012 13:29 GMT


CDN SERVER CLUSTER
LOAD BALANCER
CDN SERVER CLUSTER
LOAD BALANCER
REQUEST ROUTER
CLIENT




Map of Hydrocarbon Permits as of December 2008 Created by Superimposing UNMIG Titles onto Google Maps
Extraction Permit Awarded
Extraction Permit Pending
Exploration Permit Pending
Exploration Permit Awarded
Storage
WWF Abruzzo

Before You Attended the Course

This machine learning stuff looks cool!

Before You Attended the Course

This machine learning stuff looks cool!

Maybe I can even find a way to make my computer to learn how to write my PhD thesis...

During the Course

Theorem

Let Z_n be a training set of n i.i.d. samples from a distribution $\mathcal{P}$ and $\mathcal{H}$ be a hypothesis space with $VC(\mathcal{H}) = d$. If

$$\hat{h}(\cdot; Z_n) = \arg\min_{h \in \mathcal{H}} \widehat{R}(h; Z_n) \qquad h^*(\cdot; \mathcal{P}) = \arg\min_{h \in \mathcal{H}} R(h; \mathcal{P})$$

then

$$R(\hat{h}; \mathcal{P}) \leq R(h^*; \mathcal{P}) + O\Big(\sqrt{\frac{d \log n/\delta}{n}}\Big)$$

with probability at least $1 - \delta$ (w.r.t. the randomness in the training set).

During the Course

Theorem

Let Z_n be a training set of n i.i.d. samples from a distribution $\mathcal{P}$ and $\mathcal{H}$ be a hypothesis space with $VC(\mathcal{H}) = d$. If

$$\hat{h}(\cdot; Z_n) = \arg\min_{h \in \mathcal{H}} \widehat{R}(h; Z_n) \qquad h^*(\cdot; \mathcal{P}) = \arg\min_{h \in \mathcal{H}} R(h; \mathcal{P})$$

then

$$R(\hat{h}; \mathcal{P}) \leq R(h^*; \mathcal{P}) + O\Big(\sqrt{\frac{d \log n/\delta}{n}}\Big)$$

with probability at least $1 - \delta$ (w.r.t. the randomness in the training set).

Theorem

If $\mathcal{D}$ is a convex decision space and the loss function is bounded and convex in the first argument, then on any sequence $\mathbf{y}^n$, EWA(η) satisfies

$$R_n = L_n(\mathcal{A}; \mathbf{y}^n) - \min_i L_{i,n}(\mathbf{y}^n) \leq \frac{\log N}{\eta} + \frac{\eta n}{8}.$$

During the Course

Theorem

Let Z_n be a training set of n i.i.d. samples from a distribution $\mathcal{P}$ and $\mathcal{H}$ be a hypothesis space with $VC(\mathcal{H}) = d$. If

$$\hat{h}(\cdot; Z_n) = \arg\min_{h\in\mathcal{H}} \widehat{R}(h; Z_n) \qquad h^*(\cdot; \mathcal{P}) = \arg\min_{h\in\mathcal{H}} R(h; \mathcal{P})$$

then

$$R(\hat{h}; \mathcal{P}) \leq R(h^*; \mathcal{P}) + O\Big(\sqrt{\frac{d \log n/\delta}{n}}\Big)$$

with probability at least $1 - \delta$ (w.r.t. the randomness in the training set).

Theorem

If $\mathcal{D}$ is a convex decision space and the loss function is bounded and convex in the first argument, then on any sequence $\mathbf{y}^n$, EWA(η) satisfies

$$R_n = L_n(\mathcal{A}; \mathbf{y}^n) - \min_i L_{i,n}(\mathbf{y}^n) \leq \frac{\log N}{\eta} + \frac{\eta n}{8}.$$

Theorem

For any set of N arms with distributions bounded in $[0, b]$, if $\delta = 1/t$, then UCB(ρ) with $\rho > 1$, achieves a regret

$$R_n(\mathcal{A}) \leq \sum_{i\neq i^*} \left[\frac{4b^2}{\Delta_i}\rho\log(n) + \Delta_i\Big(\frac{3}{2} + \frac{1}{2(\rho-1)}\Big)\right]$$

After You Attended the Course

This machine learning stuff looks cool!

After You Attended the Course

This machine learning stuff looks ~~cool~~!

After You Attended the Course

This machine learning stuff looks ***awesome***!

Thank you!!

Alessandro Lazaric
alessandro.lazaric@inria.fr
sequel.lille.inria.fr